**Introduction**

For this project I chose the problem of where to build a new grocery distribution center within the given Toronto borough of Scarborough. The description is as follows:

There is a groceries distribution center in the Scarborough borough of Toronto. The main service of this distribution center is to provide the local establishments with fresh and high-quality groceries. The distribution center wants to build a warehouse for the groceries it buys from farmers inside the borough so that they will be better equipped to provide the area with better/faster service and better-quality goods.

The distribution center should build this warehouse where it is closest to its customers in order to minimize the cost of transportation. Which neighborhood (in the Scarborough borough) would be a better choice for the distribution center to build the warehouse? Finding the right neighborhood is our goal!

This type of problem has many valuable lessons that apply to many different types of industry. Amazon, for example has been doing something very similar in their search for a new east coast distribution center. Any business that is looking for a new location would find this type of analysis very handy in determining their future building site.

**Data**

We will need geo-locational information about the Scarborough borough and the neighborhoods in that borough. We will first find neighborhoods inside Scarborough by their corresponding Postal Codes. Then we will need data about different venues in different neighborhoods of that specific borough. In order to gain the desired information. We will use Foursquare locational information. As basic information, we can obtain its precise latitude and longitude and also its distance from the center of the neighborhood. But we are looking for advanced information such as the category of that venue and whether this venue is a popular one in its category or maybe the average price of the services of this venue.

**Methodology**

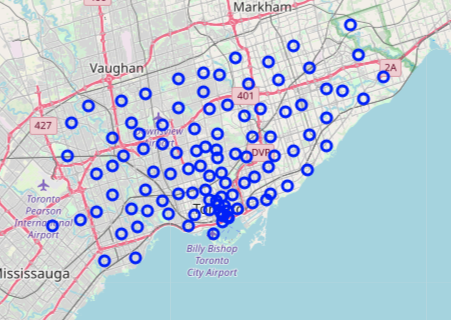
The first step in determining the solution to the given problem of what neighborhood should the client build the new grocery distribution center is to import our needed libraries, data about Toronto neighborhoods, and then clean the data. Here is an output of our cleaned data that contains information related to the neighborhoods that are found within the Scarborough borough.

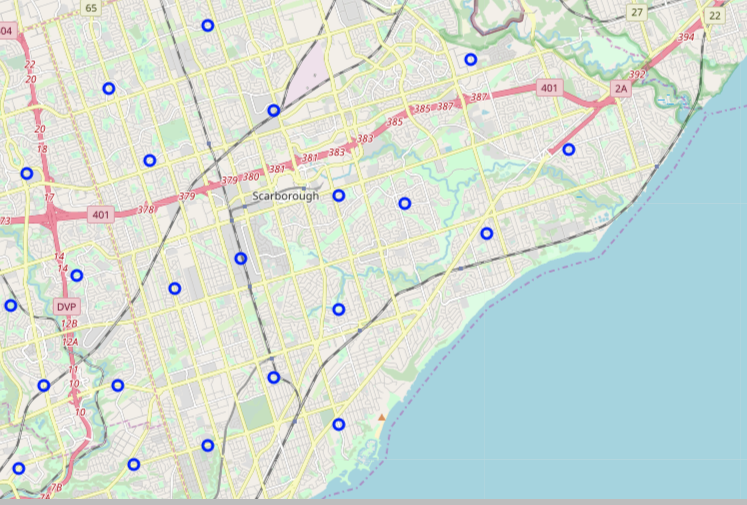


Now that we have the neighborhood information, we will move on to getting the geo-spacial (Latitude and Longitude) information for the given neighborhoods.



Next, we will generate a folium map to show the neighborhoods within Toronto and more specifically the Scarborough borough:

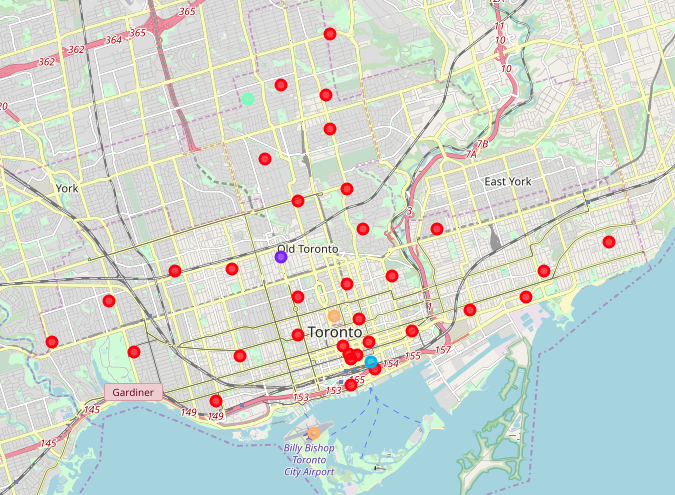




Our next step is to use the 4 Square API to determine what types of business establishments are in each neighborhood. This will help determine out potential clientele for our new distribution center and we will determine the frequency of each business within each specific neighborhood. By doing this we can start to form a picture of where we think the best place for our new distribution center will be. As we can see the top ten venues per neighborhood are as follows:

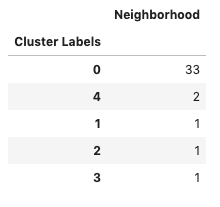


Our next step will be to use K-Clustering to determine, more in depth, where the best potential site will be located.



As we start looking further in-depth, we start to realize that the clustering of our potential clientele base is around the East Toronto/Central Toronto Boroughs.





**Results**

As we can tell based off of our clustering, there is a huge majority of potential clients that reside within the Zero clustering group which includes the East Toronto/Central Toronto Boroughs. Based off of this information, I would determine that the closer we can build within the Scarborough borough to the East Toronto/Central Toronto boroughs, the better we would be able to service our potential clients. Our next step, in my opinion, would be to determine land cost, construction costs, ect so that we can then base our distribution center within the best place to minimize cost while maximizing potential profits gained

**Discussion**

I feel the next step would be to determine land cost, construction costs, ect so that we can then base our distribution center within the best place to minimize cost while maximizing potential profits gained. As previously mentioned, the best location would be where our new distribution center would have the best access to our potential clientele and as we can see, the majority of that potential client base resides in the East Toronto/Central Toronto boroughs.

**Conclusion**

All of our python code was successfully run and provided valuable information in determining a potential solution to our given problem of where we should build our new distribution center. In going through different methods, I determined that K-Clustering was a suitable method for helping determine the best location. After importing venue information from 4 Square we generated maps and data frames of our potential cliental and where the businesses were located. This in turn helps us determine where we would like to build our new distribution center. Since part of our problem was to provide the best value of our products, we would like to minimize the time it would take to acquire our goods from farmers, for example, and then distribute those goods to our clients. With this in mind we want to look at those locations that would help us provide the shortest turnaround of our goods to help maximize freshness as well as value.